Amendments to the Specification

Please replace the paragraph at page 13, lines 14 through 18 with the following amended paragraph:

At t_7, the event monitor sees a save operation to Document E, and at time t_8, the event monitor 180 reports event 208: Document E is sent over the Internet. Has the user stored and sent information from a sensitive Document A" as Document E, compromising security? Or [[is]] has she just created a birthday invitation Document E from Document D?

Please replace the paragraph at page 17, line 25 through page 18, line 4 with the following amended paragraph:

The clusters within the hierarchy have a substantial degree of intersection, so that going down all the branches of the tree where the similar clusters might possibly be found drives the query down most of the branches and eliminates the benefit of having a hierarchy (as compared to a simple set of clusters). The query uses probabilistic estimates to determine which clusters are most likely the hosts of the given chunk and proceeds only to explore the ~~braches~~ branches of the hierarchy, passing through these clusters. This multi-branch, probabilistic search provides a configurable balance between the required accuracy and performance that is vital to determine document similarity in real time.

Please replace the paragraph at page 23, lines 12 through 16 with the following amended paragraph:

We chose to construct an algorithm similar to the "Growing Hierarchical Self-Organizing Maps" method described in Dittenbach, M., Rauber, A., and Merkl, D., "Uncovering the Hierarchical Structure in Data Using the Growing Hierarchical Self-Organizing Map", *Neurocomputing*, 2002, 48(1-4):199-216~~, http://www.ifs.tuwien.ac.at/~mbach/ghsom/~~.